

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE  
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES**

In re Application of:	)	
	)	
<b>Gregory J. Lauckhart, et al.</b>	)	
	)	
Serial No.: 09/695,216	)	Group Art Unit: 2142
	)	
Filed: July 24, 2002	)	Examiner: John Moore Frink
	)	
For: SYSTEM AND METHOD	)	
ESTIMATING PREVALENCE	)	
OF DIGITAL CONTENT ON	)	
THE WORLD-WIDE-WEB	)	

Mail Stop: Appeal Brief - Patents  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

**APPEAL BRIEF UNDER 37 C.F.R. § 41.37**

In support of the Notice of Appeal filed on May 13, 2008, and pursuant to 37 C.F.R. § 41.37, Appellants present this appeal brief in the above-captioned application.

This is an appeal to the Board of Patent Appeals and Interferences from the Examiner's final rejection of claims 1-69 in the Final Office Action dated November 13, 2007. The appealed claims are set forth in the attached Claims Appendix.

1. Real Party in Interest

This application was originally assigned to Jupiter Media Metrix, Inc. and was then assigned to NetRatings, Inc., which is the real party in interest.

2. Related Appeals and Interferences

There are no other appeals or interferences which would directly affect, be directly affected by, or have a bearing on the instant appeal.

3. Status of Claims

Claims 1-69 have been rejected in the final Office Action. For clarity, the specific claim rejections are listed in the following format:

- Claims 1-2, 4-10, 14, 15, 18-21, 23, 25, 32, 35,55, 56, 58-61, 63-66, 68 and 69 stand rejected under 35 U.S.C. §103(a) as being unpatentable over U.S. Patent No. 5,995,943 (“Bull”) in view U.S. Patent No. 5,784,635 (“McCallum”) in view of U.S. Patent No. 5,878,426 (“Plasek”).
- Claims 3, 22, 24, 26, 29 and 33 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Bull in view of McCallum in view of Plasek in view of U.S. Patent No. 6,601,100 (“Lee”).
- Claims 11 and 27 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Bull in view of McCallum in view of Plasek in view of “Consistent, Yet Anonymous, Web Access with LPWA, Feb. 1999” (“Gabber”).
- Claims 12, 13 and 28 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Bull in view of McCallum in view of Plasek in view of Lee.
- Claims 30 and 31 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Bull in view of McCallum in view of Plasek in view of Lee in view of U.S. Patent No. 5,878,213 (“Bittinger”).
- Claim 34 stands rejected under 35 U.S.C. §103(a) as being unpatentable over Bull in view of McCallum in view of Plasek in view of U.S. Patent No. 6,167,402 (“Yeager”).

- Claim 36 stands rejected under 35 U.S.C. §103(a) as being unpatentable over Bull in view of McCallum in view of Plasek in view of U.S. Patent No. 6,094,684 (“Pallmann”).
  - Claims 16, 17, 37, 44, 48, 50, 53 and 54 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Bull in view of McCallum in view of Plasek in view of 6,108,637 (“Blumenau”).
  - Claims 38, 41, 45, 49 and 51 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Bull in view of McCallum in view of Plasck in view of Blumenau in view of Lee.
  - Claims 39, 40 and 52 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Bull in view of McCallum in view of Plasek in view of Blumenau in view of Lee in view of Gabber.
  - Claims 42 and 43 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Bull in view of McCallum in view of Plasek in view of Blumenau in view of Lee in view of Bittinger.
  - Claim 37 stands rejected under 35 U.S.C. §103(a) as being unpatentable over Bull in view of McCallum in view of Plasek in view of Blumenau in view of Yeager.
  - Claim 46 stands rejected under 35 U.S.C. §103(a) as being unpatentable over Bull in view of McCallum in view of Plasek in view of Blumenau in view of Pallmann.
- The final rejection of claims 1-69 is being appealed.

4. Status of Amendments

All amendments submitted by the Appellants have been entered.

5. Summary of Claimed Subject Matter

Claim 1 recites a system for estimating the prevalence of digital content on a network, (see, e.g. Fig. 1 and ¶ 29). The system includes an estimating device that receives traffic data collected from the network, (see, e.g. Traffic Analysis System 210, ¶ 34, lines 1-2), an anonymizing device that locates user identification data in the traffic data, masks the user identification data to produce clean traffic data, and stores the clean

traffic data, (see, e.g., Anonymity System 310, Fig. 3, ¶ 40), a sampling device that stores summarization data that describes each occurrence of the digital content in the clean traffic data and scales the data by a weighting factor to extrapolate global traffic data, (see, e.g. Traffic Summarization 312, Fig. 3, ¶ 41), and an accessing device that presents the clean traffic data and the summarization data to a user, (see, e.g. User Interface 240, ¶ 33, lines 5-6).

Claim 6 recites a method of estimating the prevalence of digital content on a network, (see, e.g. Fig. 1 and ¶ 29). The method includes estimating the global traffic to at least one Web site on the network to provide traffic data, (see, e.g. Traffic Analysis System 210, ¶ 34, lines 1-2), locating user identification data in the traffic data and masking the user identification data to produce clean traffic data, (see, e.g., Anonymity System 310, Fig. 3, ¶ 40), statistically sampling the contents of said at least one Web site to provide sampling data including scaling the data by a weighting factor to extrapolate global traffic data and storing the clean traffic data and the sampling data, (see, e.g., Traffic Summarization 312, Fig. 3, ¶ 41), and accessing the clean traffic data and the sampling data to generate a report, (see, e.g., User Interface 240, ¶ 33, lines 5-6).

Claim 7 recites a system for estimating the prevalence of digital content on a network (see, e.g. Fig. 1 and ¶ 29) connected to at least one network site that includes at least one network server to access at least one uniform resource locator (see, e.g. Fig. 1, 110, 112). The system includes a database (see, e.g. database 200), a traffic analysis system that receives traffic data from a traffic sampling system (see, e.g., 210, ¶ 340, lines 1-2), locates user identification data in the traffic data, masks the user identification data to produce clean traffic data, and stores the clean traffic data in the database (see,

e.g. 310, Fig. 3, ¶ 41), the traffic data including said at least one uniform resource locator. The system including a digital content sampling system that stores the digital content at said at least one uniform resource locator in the database (see, e.g., 220, Fig. 2, ¶ 35) and a statistical summarization system that stores summarization data that describe the digital content in the database including scaling the data by a weighting factor to extrapolate global traffic data (see, e.g. 230, Fig. 2, ¶ 36).

Claim 25 recites system for estimating prevalence of digital content on a network including a memory device and a processor disposed in communication with the memory device (see, e.g. Fig. 1, ¶ 8, ¶ 29). The processor is configured to obtain traffic data from at least one Web site on the network (see, e.g., ¶ 34, lines 1-2, Fig. 2, element 210), locate user identification data in the traffic data and mask the user identification data to produce clean traffic data (see, e.g., Fig. 3, ¶ 40, element 310), compute a number of impressions for the digital content in the clean traffic data including scaling the number by a weighting factor to extrapolate global traffic data and retrieve the digital content from the clean traffic data to generate sampling data (see, e.g., Fig. 3, ¶ 41, element 312) and generate prevalence estimates for the digital content from the clean traffic data and the sampling data (see, e.g., ¶ 57, see also Figs. 4A-4C and ¶¶ 62-66).

Claim 37 recites a method for using a computer to estimate the prevalence of digital content on a network (see, e.g. Fig. 1, ¶ 8, ¶ 29). The method includes obtaining traffic data from at least one Web site on the network (see, e.g. ¶ 34, lines 1-12, Fig. 2, element 210), locating user identification data in the traffic data and masking the user identification data to produce clean traffic data, (see, e.g., Fig. 3, ¶ 40, element 310), computing a number of impressions for the digital content in the clean traffic data

including scaling the data by a weighting factor to extrapolate global traffic data and retrieving the digital content from the clean traffic data to generate sampling data (see, e.g., Fig. 3, ¶ 41, element 312), and generating prevalence estimates for the digital content from the clean traffic data and the sampling data (see, e.g., ¶ 57, see also Figs. 4A-4C and ¶¶ 62-66).

Claim 48 recites a computer readable medium (see, e.g. ¶ 8) including code for obtaining traffic data from at least one Web site on the network (see, e.g., ¶ 34, lines 1-2, Fig. 2, element 210), code for locating user identification data in the traffic data and code for masking the user identification data to produce clean traffic data (see, e.g., Fig. 3, ¶ 40, element 310), code for computing a number of impressions of digital content in the clean traffic data including scaling the data by a weighting factor to extrapolate global traffic data and code for retrieving the digital content from the clean traffic data to generate sampling data (see, e.g., Fig. 3, ¶ 41, element 312) and code for generating prevalence estimates for the digital content from the clean traffic data and the sampling data (see, e.g., ¶ 57, see also Figs. 4A-4C and ¶¶ 62-69).

Claim 50 recites a system for estimating prevalence of digital content on a network (see, e.g., Fig. 1, ¶ 8, ¶ 29). The system includes means for obtaining traffic data from at least one Web site on the network (see, e.g., ¶ 34, lines 1-2, Fig. 2, element 210), means for locating user identification data in the traffic data and means for masking the user identification data to produce clean traffic data (see, e.g., Fig. 3, ¶ 40, element 310), means for computing a number of impressions for the digital content in the clean traffic data including scaling the data by a weighting factor to extrapolate global traffic data and means for retrieving the digital content from the clean traffic data to generate sampling

data (see, e.g., Fig. 4, ¶ 41, element 312) and means for generating prevalence estimates of the digital content from the clean traffic data and the sampling data (see, e.g. ¶ 57, see also Figs. 4A-4C and ¶¶ 62-69).

Claim 53 recites a system of estimating prevalence of digital content on a network (see, e.g. Fig. 1, ¶ 29). The system includes means for estimating global traffic to at least one Web site on the network to provide traffic data (see, e.g., Fig. 2, element 210, ¶ 34), means for locating user identification data in the traffic data and means for masking the user identification data to produce clean traffic data (see, e.g., Fig. 3, ¶ 40, element 310), means for statistically sampling the contents of said at least one Web site to provide sampling data including scaling the data by a weighting factor to extrapolate global traffic data and means for storing the clean traffic data and the sampling data (see, e.g., Fig. 3, ¶ 41, element 312) and means for generating prevalence estimates for the digital content by accessing the clean traffic data and the sampling data (see, e.g., ¶ 57, see also Figs. 4A-4C and ¶¶ 62-66).

Claim 55 recites a method for using a computer to estimate the prevalence of digital content on a network (see, e.g. Fig. 1, ¶ 29). The method includes receiving traffic data from the network (see, e.g. ¶ 34, Fig. 2, element 210), locating user identification data in the traffic data and masking the user identification data to produce clean traffic data (see, e.g., ¶ 40, Fig. 3, element 310), storing the clean traffic data and storing summarization data that describe each occurrence of the digital content in the clean traffic data, where the summarization data includes data scaled by a weighting factor to extrapolate global traffic data (see, e.g., Fig. 3, ¶ 41, element 312) and presenting the

clean traffic data and the summarization data to a user (see, e.g., ¶ 57, see also Figs. 4A-4C and ¶¶ 62-66).

Claim 60 recites a system for estimating prevalence of digital content on a network(see, e.g. Fig. 1, ¶ 29). The system includes a memory device; and a processor disposed in communication with the memory device, the processor configured to receive traffic data from the network (see, e.g., ¶ 34, Fig. 2, element 210), locate user identification data in the traffic data and mask the user identification data to produce clean traffic data (see, e.g., Fig. 3, ¶ 40, element 310), store the clean traffic data and store summarization data that describe each occurrence of the digital content in the clean traffic data, where the summarization data includes data scaled by a weighting factor to extrapolate global traffic data (see, e.g., Fig. 3, ¶ 41, element 312) and present the clean traffic data and the summarization data to a user (see, e.g., ¶ 57, see also Figs. 4A-4C and ¶¶ 62-66).

Claim 65 recites a computer readable medium (see, e.g., ¶ 8, ¶ 29) comprising code for receiving traffic data from the network (see, e.g., ¶ 34, Fig. 2, element 210) including code for locating user identification data in the traffic data, code for masking the user identification data to produce clean traffic data (see, e.g., Fig. 3, ¶ 40, element 310) and code for storing the clean traffic data and code for storing summarization data that describe each occurrence of the digital content in the clean traffic data, where the summarization data includes data scaled by a weighting factor to extrapolate global traffic data (see, e.g. Fig. 3, ¶ 41, element 312) and code for presenting the clean traffic data and the summarization data to a user(see, e.g., ¶ 34, Fig. 2, element 210).



6. Grounds of Rejection to be Reviewed on Appeal

I. The Examiner improperly rejects claims 1, 7, 25, 37, 48, 50, 53, 55, 60 and 65 under 35 U.S.C. §103(a) because U.S. Patent No. 5,995,943 (“Bull”) fails to teach or suggest the claimed “traffic data.”

7. Argument

**I. The Examiner improperly rejects claims 1, 7, 25, 37, 48, 50, 53, 55, 60 and 65 under 35 U.S.C. §103(a) because Bull fails to teach or suggest the claimed “traffic data.”**

Appellants respectfully submit that all claims, including Independent claims 1, 7, 25, 37, 48, 50, 53, 55, 60 and 65 stand improperly rejected under 35 U.S.C. §103(a) based at least on the combination of Bull in view of McCallum in view of Plasek. It is noted that claims 37, 48 and 50 stand rejected in combination with additional prior art references to the Bull, McCallum, Plasek combination.

The rejections are improper because Bull fails to teach or suggest the claimed “traffic data.” For example, claim 1 recites an “estimating device that receives traffic data collected from the network.” In supporting this rejection, the Examiner asserts that Bull describes the claimed traffic data collected from the network, to which Appellants respectfully disagree.

Bull describes a Datastore that stores data relating to an individual user’s activity to enhance “URL Munging” to the specific user, e.g. providing as many personalized hyperlinks or advertisements to the user. More specifically, Bull tracks where within an existing computing system a user requested information. In the examples in the Bull

system, this tracking includes noting the contextual information of which towns the user requested information on, which cities the user inquired about airfare, did the user inquire about a particular sporting event. In the Bull system, this data is collected and tracked based on content, as is expressly described in col. 3, lines 51-59 which describes the information being “topically orientated” using the HyperText Mark-up Language content. Stated more directly, Bull looks at the content of the user’s activities, not the traffic, and hence does not teach or suggest the traffic data as claimed.

In the Response to Arguments section of the Office Action dated November 13, 2007, beginning on page 19, the Examiner asserts further citations to Bull, to which Appellants respectfully submit fail to support the assertion of Bull teaching or suggesting the claimed “traffic data,” but rather these passages, confirm the teaching or suggesting of content, which is inconsistent with traffic data.

The Examiner cites to col. 3, lines 52-54, which as Appellants noted above explicitly describes the contextual data of Bull, not the claimed traffic data. The Examiner cites to col. 5, lines 51-54, which again is contextual as this passage describes the users viewing habits. The contextual nature of Bull is further evidenced by preceding passages that describe using a “text analysis tool” to search the text of the web pages being visited, see col. 5, lines 36-40.

The Examiner cites to col. 5, lines 62-64, which again supports the contextual nature of the Bull system, where the contextual data is not traffic data. In Bull system, once a user performs a search, the same contextual operations are utilized to determine if there is a market or need being unmet by the search results. Bull gives the example of the context of the search being snorkeling off the coast of Texas and after a 100 searches,

selling this information to a tour provider. This is contextual, not traffic data. Bull does not teach or suggest tracking user traffic information once a search is completed in this cited passage, it merely notes the contextual information that 100 users have performed a search for the same thing.

The Examiner also cites to col. 7, lines 45-46, which again describe a contextual embodiment, which in this embodiment relates not even to the context of user activities, but rather is an optimization technique for low bandwidth applications. This passage merely describes the common technique for replacing bandwidth consuming files with tags to improve transmission speeds for low-bandwidth users. This is completely unrelated to traffic data.

The Examiner also cites to col. 8, lines 29-31, which describes the storage of what the person is viewing. Again, this relates to contextual operations. This passage describes lead generation operations such that the Bull system can sell user information as possible leads to people who wish to conduct an advertising campaign. In the lines preceding the Examiner-cited passage, col. 8, lines 26-27 clearly and explicitly state that the lead generation uses “complex software text search agents” to performing the searching of user activities and hence generate the leads. This is entirely and absolutely consistent with the contextual nature of Bull, wherein contextual information is inconsistent with the claimed traffic information.

To re-iterate, Bull describes a system that tracks the type of content that the user views, such as the examples of “Germany travel, the Olympics, Spring Break or even new cars.” (col. 3, lines 51-52). The words that describe the user-visited content is then used to link corresponding advertisements, among other features. The detection and

logging of the type of content that a user accesses is wholly inconsistent with the claimed “traffic data.” Traffic data, by its plain and ordinary meaning relates to data about traffic and not data about the content of the traffic. This distinction defines a fundamental difference between the claimed invention and the Examiner’s application of Bull, where Bull relates to determining the types of content and context-based activities that a user performed online, the claimed “traffic data” concerns itself with data relating to the traffic across the network.

Moreover, while the Examiner is given a broad and reasonable scope in interpreting claims during examination, Appellants submit that it is overly broad and hence unreasonable to find Bull’s content data as teaching or suggesting the claimed traffic data.

Appellants submit that all dependent claims are similarly patentable in view of the shortcomings of the combination of at least Bull, McCallum and Plasek, for at least the reasons stated above. Accordingly, Appellants submit that all pending claims are patentable and the present rejections are thereby improper.

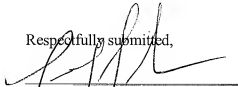
8. Conclusions

For the reasons set forth above, Appellants respectfully request that the Board reverse the final rejections of the claims by the Examiner under 35 U.S.C §103(a) and indicate that claims 1-69 are allowable.

Dated: July 14, 2008

THIS CORRESPONDENCE IS BEING SUBMITTED  
ELECTRONICALLY THROUGH THE PATENT AND  
TRADEMARK OFFICE EFS FILING SYSTEM ON  
July 14, 2008.

Respectfully submitted,



Timothy J. Bechen

Reg. No. 48,126

DREIER LLP

499 Park Ave.

New York, New York 10022

Tel : (212) 328-6100

Fax: (212) 328-6101

***Customer No. 61834***

**Claims Appendix**

1. A system for estimating the prevalence of digital content on a network, comprising:
  - an estimating device that receives traffic data collected from the network;
  - an anonymizing device that locates user identification data in the traffic data, masks the user identification data to produce clean traffic data, and stores the clean traffic data;
  - a sampling device that stores summarization data that describes each occurrence of the digital content in the clean traffic data and scales the data by a weighting factor to extrapolate global traffic data; and
  - an accessing device that presents the clean traffic data and the summarization data to a user.
2. The system of claim 1, wherein the estimating device receives the traffic data from at least one proxy cache server.
3. The system of claim 1, wherein the sampling device computes the number of impressions of the digital content for a web site on the network.
4. The system of claim 1, wherein the sampling device includes:
  - a prober that fetches a web page from the network;
  - an extractor that locates a fragment of the web page that includes the digital content; and
  - a classifier that performs a structural analysis of the fragment to classify the digital content.

5. The system of claim 1, wherein the accessing device generates a report when the clean traffic data or the summarization data satisfy at least one criterion.

6. A method of estimating the prevalence of digital content on a network, comprising the steps of:

estimating the global traffic to at least one Web site on the network to provide traffic data;

locating user identification data in the traffic data;

masking the user identification data to produce clean traffic data;

statistically sampling the contents of said at least one Web site to provide sampling data including scaling the data by a weighting factor to extrapolate global traffic data;

storing the clean traffic data and the sampling data; and

accessing the clean traffic data and the sampling data to generate a report.

7. A system for estimating the prevalence of digital content on a network connected to at least one network site that includes at least one network server to access at least one uniform resource locator, the system comprising:

a database;

a traffic analysis system that receives traffic data from a traffic sampling system, locates user identification data in the traffic data, masks the user identification data to produce clean traffic data, and stores the clean traffic data in the database, the traffic data including said at least one uniform resource locator;

a digital content sampling system that stores the digital content at said at least one uniform resource locator in the database; and

a statistical summarization system that stores summarization data that describe the digital content in the database including scaling the data by a weighting factor to extrapolate global traffic data.

8. The system of claim 7, further comprising:

a Web front end that connects to the network and the database, wherein a client uses a browser to connect to the Web front end.

9. The system of claim 7, further comprising:

a user interface that an account manager, an operator, or a media editor can use to administer the system.

10. The system of claim 7, wherein the network is the Internet, and

wherein the network site is a Web site.

11. The system of claim 7, wherein to mask the user identification data

in the traffic data the traffic analysis system replaces the user identification data with a result from processing the user identification data through a cryptographically secure one-way hash function.

12. The system of claim 11, wherein the user identification data

includes a network address or a cookie.

13. The system of claim 11, wherein the summarization data includes a

reference to said at least one uniform resource locator and a count of the number of requests for said at least one uniform resource locator.



14. The system of claim 7, wherein the digital content sampling system further comprises:
- a probemapping system that uses the summarization data to create a probe map for the network, the probe map including a mapping for said at least one uniform resource locator;
  - a uniform resource locator retrieval system that retrieves said at least one uniform resource locator from the network server;
  - a browser emulation environment that conducts a simulation of the display of said at least one uniform resource locator in a browser;
  - a digital content extractor that stores the digital content from said at least one uniform resource locator in the database; and
  - a structural classifier that stores at least one classification type for the digital content in the database.
15. The system of claim 14, wherein the probe map further comprises:
- a probability that said at least one uniform resource location will be sampled; and
  - a scale that determines the contribution of said at least one uniform resource location to the summarization data.
16. The system of claim 14, wherein the simulation includes executing a program referenced by said at least one uniform resource locator.
17. The system of claim 16, wherein the program is a JavaScript script, a Java applet, a Perl script, or a common gateway interface program.

18. The system of claim 14, wherein the simulation includes executing dynamic digital content referenced by said at least one uniform resource locator.

19. The system of claim 18, wherein the dynamic content is an interlaced GIF image, an MPEG movie, or an MP3 audio file.

20. The system of claim 14, wherein the digital content extractor retrieves the digital content from a location designated by said at least one uniform resource locator by applying a rule set defined by a media editor.

21. The system of claim 14, wherein the digital content extractor retrieves the digital content from a location designated by said at least one uniform resource locator by using an automated digital content detection system.

22. The system of claim 21, wherein the automatic digital detection system comprises:

a structural detector that locates an XML structure; and

a feature detector that locates an XML feature within the XML structure.

23. The system of claim 14, wherein the structural classifier determines said at least one classification type for the digital content.

24. The system of claim 7, wherein the user interface further comprises:

a system account management interface that assists an account manager with creating and modifying an account on the system;

a site administration interface that assists the operator with the administration of said at least one network site;

a taxonomy administration interface that assists the media editor with the administration of taxonomy data;

a digital content classification interface that assists the media editor with the classification of the digital content; and

a rate card collection interface that assists the media editor with the administration of rate card data.

25. A system for estimating prevalence of digital content on a network, comprising:

a memory device; and

a processor disposed in communication with the memory device, the processor configured to:

obtain traffic data from at least one Web site on the network;

locate user identification data in the traffic data;

mask the user identification data to produce clean traffic data;

compute a number of impressions for the digital content in the clean traffic data including scaling the number by a weighting factor to extrapolate global traffic data;

retrieve the digital content from the clean traffic data to generate sampling data; and

generate prevalence estimates for the digital content from the clean traffic data and the sampling data.

26. The system of claim 25, wherein the processor is further configured to:

retrieve a Web page from said at least one Web site;  
extract a fragment from the Web page; and  
classify the fragment.

27. The system of claim 25, wherein to mask the user identification data in the traffic data the processor is further configured to:

replace the user identification data with a result from processing the user identification data through a cryptographically secure one-way hash function.

28. The system of claim 27, wherein the user identification includes a network address or a cookie.

29. The system of claim 25, wherein the processor is further configured to:

classify a fragment within the sampling data.

30. The system of claim 29, wherein the processor is further configured to:

classify the fragment by analyzing the fragment for uniqueness and adding information to a database regarding the uniqueness of the fragment.

31. The system of claim 30, wherein the processor is configured to:  
classify the fragment by detecting a duplicate fragment.

32. The system of claim 25, wherein the processor is further configured to:

interact with a user interface that administers the system.

33. The system of claim 25, wherein the processor is further configured to:

include uniform resource locator information regarding said at least one Web site in the traffic data.

34. The system of claim 25, wherein the processor is further configured to:

perform data integrity monitoring of the sampling data.

35. The system of claim 25, wherein the processor is further configured to:

serve as an automatic digital content detection system.

36. The system of claim 35, wherein the automatic advertisement detection system applies at least one heuristic algorithm to detect digital content within an HTML or an XML document and normalizes the detected HTML or XML content into a hierarchical form.

37. A method for using a computer to estimate the prevalence of digital content on a network, comprising the steps of:

obtaining traffic data from at least one Web site on the network;  
locating user identification data in the traffic data;  
masking the user identification data to produce clean traffic data;  
computing a number of impressions for the digital content in the clean traffic data including scaling the data by a weighting factor to extrapolate global traffic data;

retrieving the digital content from the clean traffic data to generate sampling data; and

generating prevalence estimates for the digital content from the clean traffic data and the sampling data.

38. The method of claim 37, wherein retrieving the digital content further comprises the steps of:

retrieving a Web page from said at least one Web site;

extracting a fragment from the Web page; and

classifying the fragment.

39. The method of claim 37, wherein the masking of the user identification data in the traffic data further comprises:

replacing the user identification data with a result from processing the user identification data through a cryptographically secure one-way hash function.

40. The method of claim 39, wherein the user identification includes a network address or a cookie.

41. The method of claim 37, further comprising the classifying a fragment within the sampling data.

42. The method of claim 41, wherein classifying the fragment further comprises the steps of:

analyzing fragment for uniqueness; and

adding information to a database regarding the uniqueness of the fragment.

43. The method of claim 42, further comprising the step of:

classifying the fragment by detecting a duplicate fragment.

44. The method of claim 37, further comprising the step of:  
interacting with a user interface that administers the system.

45. The method of claim 37, further comprising the step of:  
including uniform resource locator information regarding said at

least one Web site in the traffic data.

46. The method of claim 37, further comprising the step of:  
performing data integrity monitoring of the sampling data.

47. The method of claim 37, further comprising the steps of:  
performing automatic advertisement detection by applying at least  
one heuristic algorithm to detect advertising within an HTML or an XML document; and  
normalizing the detected HTML or XML content into a  
hierarchical form.

48. A computer readable medium comprising:  
code for obtaining traffic data from at least one Web site on the  
network;  
  
code for locating user identification data in the traffic data;  
  
code for masking the user identification data to produce clean  
traffic data;  
  
code for computing a number of impressions of digital content in  
the clean traffic data including scaling the data by a weighting factor to extrapolate global  
traffic data;

code for retrieving the digital content from the clean traffic data to generate sampling data; and

code for generating prevalence estimates for the digital content from the clean traffic data and the sampling data.

49. The computer readable medium of claim 48, further comprising:

code for retrieving a Web page from said at least one Web site; code for extracting a fragment from the Web page; and code to classify the fragment.

50. A system for estimating prevalence of digital content on a network, comprising:

means for obtaining traffic data from at least one Web site on the network;

means for locating user identification data in the traffic data;

means for masking the user identification data to produce clean traffic data:

means for computing a number of impressions for the digital content in the clean traffic data including scaling the data by a weighting factor to extrapolate global traffic data;

means for retrieving the digital content from the clean traffic data to generate sampling data; and

means for generating prevalence estimates of the digital content from the clean traffic data and the sampling data.

51. The system of claim 50, further comprising:

means for classifying a fragment extracted from a Web page.



52. The system of claim 50, further comprising:

means for replacing the user identification data with a result from processing the user identification data through a cryptographically secure one-way hash function.

53. A system of estimating prevalence of digital content on a network, comprising:

means for estimating global traffic to at least one Web site on the network to provide traffic data;

means for locating user identification data in the traffic data;

means for masking the user identification data to produce clean traffic data;

means for statistically sampling the contents of said at least one Web site to provide sampling data including scaling the data by a weighting factor to extrapolate global traffic data;

means for storing the clean traffic data and the sampling data; and

means for generating prevalence estimates for the digital content by accessing the clean traffic data and the sampling data.

54. The system of claim 53, further comprising:

means for reporting the prevalence estimates to a user.

55. A method for using a computer to estimate the prevalence of digital content on a network, comprising the steps of:

receiving traffic data from the network:

locating user identification data in the traffic data;

masking the user identification data to produce clean traffic data;  
storing the clean traffic data;  
storing summarization data that describe each occurrence of the digital content in the clean traffic data, where the summarization data includes data scaled by a weighting factor to extrapolate global traffic data; and  
presenting the clean traffic data and the summarization data to a user.

56. The method of claim 55, wherein the receiving of the traffic data is from at least one proxy server.

57. The method of claim 55, wherein storing summarized traffic data further comprises the step of:

computing the number of impressions of the digital content for a web site on the network.

58. The method of claim 55, wherein storing traffic data further comprises the steps of:

fetching a web page from the network;  
locating a fragment of the web page that includes the digital content; and  
performing a structural analysis of the fragment to classify the digital content.

59. The method of claim 55, wherein presenting the clean traffic data and the summarization data further comprises the step of:

generating a report when the clean traffic data or the summarization data satisfy at least one criterion.

60. A system for estimating prevalence of digital content on a network, comprising:

a memory device; and

a processor disposed in communication with the memory device, the processor configured to:

receive traffic data from the network;

locate user identification data in the traffic data;

mask the user identification data to produce clean traffic data;

store the clean traffic data;

store summarization data that describe each occurrence of the digital content in the clean traffic data, where the summarization data includes data scaled by a weighting factor to extrapolate global traffic data; and

present the clean traffic data and the summarization data to a user.

61. The system of claim 60, wherein the receiving of the traffic data is from at least one proxy server.

62. The system of claim 60, wherein the processor computes the number of impressions of the digital content for a web site on the network.

63. The system of claim 60, wherein the processor is further configured to:

fetch a web page from the network;

locate a fragment of the web page that includes the digital content;  
and perform a structural analysis of the fragment to classify the digital content.

64. The system of claim 60, wherein the processor generates a report when the clean traffic data or the summarization data satisfy at least one criterion.

65. A computer readable medium comprising:

code for receiving traffic data from the network;

code for locating user identification data in the traffic data;

code for masking the user identification data to produce clean traffic data;

code for storing the clean traffic data;

code for storing summarization data that describe each occurrence of the digital content in the clean traffic data, where the summarization data includes data scaled by a weighting factor to extrapolate global traffic data; and

code for presenting the clean traffic data and the summarization data to a user.

66. The computer readable medium of claim 65, the receiving of the traffic data is from at least one proxy server.

67. The computer readable medium of claim 65, further comprising:

code for computing the number of impressions of the digital content for a web site on the network.

68. The computer readable medium of claim 65, further comprising:

code for fetching a web page from the network;

code for locating a fragment of the web page that includes the digital content; and code for performing a structural analysis of the fragment to classify the digital content.

69. The computer readable medium of claim 65, further comprising:  
code for generating a report when the clean traffic data or the summarization data satisfy at least one criterion.

**Evidence Appendix**

No evidence has been submitted or relied upon in the instant appeal.

**Related Proceedings Appendix**

There are no related proceedings which are related to or would have a bearing on the instant appeal.